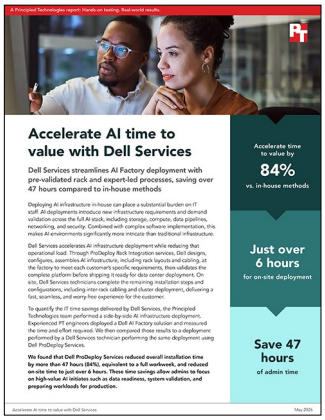

The science behind the report:

Accelerate AI time to value with Dell Services

This document describes what we tested, how we tested, and what we found. To learn how these facts translate into real-world benefits, read the report Accelerate AI time to value with Dell Services.

We concluded our hands-on testing on March 14, 2026. During testing, we determined the appropriate hardware and software configurations and applied updates as they became available. The results in this report reflect configurations that we finalized on March 9, 2026 or earlier. Unavoidably, these configurations may not represent the latest versions available when this report appears.

Our results

To learn more about how we have calculated the wins in this report, go to https://facts.pt/calculating-and-highlighting-wins. Unless we state otherwise, we have followed the rules and principles we outline in that document.

Total time

Table 1: Total time to deploy a Dell AI Factory solution using in-house engineers vs. engaging with Dell Services.

Task	Dell Services	In-house engineers
Total time	**8:11:00**	**53:15:00**
Pre-engagement	2:00:00	21:30:00
Day 1	6:11:00	6:48:00
Day 2	NA	10:02:00
Day 3	NA	11:35:00
Day 4	NA	3:20:00

Issue mitigation

Table 2: Total time to mitigate issues while deploying a Dell AI Factory solution using in-house engineers vs. engaging with Dell Services.

Task	Dell Services	In-house engineers
Total time mitigating issues	1:20:00	6:00:00
Day 1	1:20:00	NA
Day 2 (bad installer, network workarounds, documentation issues)	NA	2:00:00
Day 3 (documentation issues, benchmark container execution issues)	NA	3:00:00
Day 4 (rollback and reinstall kubernetes due to user error)	NA	1:00:00

Detailed breakdown of tasks

Table 3: Detailed breakdown of Dell AI Factory deployment tasks, by day, using in-house engineers vs. engaging with Dell Services.

Task	Dell Services	In-house engineers
Pre-engagement	**2:00:00**	**21:30:00**
Research and education (Tech 1)	NA	7:30:00
Research and education (Tech 2)	NA	4:00:00
Planning (Tech 1)	NA	4:00:00
Planning (Tech 2)	NA	4:00:00
Planning meetings	2:00:00	2:00:00
Day 1	**6:11:00**	**6:48:00**
Pre-deployment	1:00:00	3:45:00
Download image and prepare installer USB, during which we verified infrastructure addresses and inspected cabinets and cabling	NA	2:30:00
Reviewing and/or updating documentation	0:25:00	1:15:00
Worksite setup (laptops, power, etc)	0:35:00	NA
Racking and cabling	0:10:00	3:03:00
Rack compute	NA	0:25:00
Rack controllers	NA	0:13:00
Rack switches	NA	0:20:00
Cable compute with InfiniBand	NA	0:35:00
Cable controllers	NA	0:12:00
Cable switches	NA	0:18:00
Power-on test, iDRAC configuration, and MAC collection	NA	1:00:00
Cabling Dell Rack Integrated gear	0:10	NA
Head node installation and configuration	0:57:00	(Completed on Day 2)
Deploy head node	0:24:00	NA
InfiniBand configuration	0:20:00	NA
Patch/update/license	0:12:00	NA
Define compute nodes	0:01:00	NA

Task	Dell Services	In-house engineers
Initialize InfiniBand network	10:00	(Completed on Day 2)
InfiniBand configuration	0:10:00	NA
Compute node deployment (bringing up the cluster - continued)	0:50:00	(Completed on Day 2)
Node provisioning	0:50:00	NA
Logical configuration	1:09:00	(Completed on Day 2)
Install Docker	0:21:00	NA
Install Kubernetes	0:48:00	NA
Validation	0:35:00	(Completed on Day 2)
Benchmark containers (NCCL, HPL, STREAM)	0:20:00	NA
DeepSeek model	0:15:00	NA
Issue mitigation	1:20:00	(Completed on Day 2)
Issue mitigation	1:20:00	NA
Day 2	**(Dell Services work complete)**	**10:02:00**
Head node installation and configuration	NA	7:22:00
Head node installation and configuration	NA	1:55:00
Deploy head node	NA	1:15:00
Rebuild installer USB	NA	0:40:00
Bringing up the cluster	**NA**	**5:27:00**
InfiniBand configuration on head node	NA	0:55:00
Define and test compute nodes	NA	4:20:00
Patch/update/license	NA	0:12:00
Initialize InfiniBand network	NA	0:40:00
InfiniBand configuration	NA	0:40:00
Issue mitigation	**NA**	**2:00:00**
Bad installer, network workarounds, documentation issues	NA	2:00:00
Day 3	**(Dell Services work complete)**	**11:35:00**
Compute node deployment (bringing up the cluster - continued)	NA	3:45:00
Dell hardware image customization for compute nodes	NA	1:45:00
Dell hardware image customization for Kubernetes controller	NA	1:05:00
Node provisioning	NA	0:55:00

Task	Dell Services	In-house engineers
Logical configuration	NA	1:45:00
Configure Kubernetes control plane server	NA	1:15:00
Install Docker on compute nodes	NA	0:30:00
Validation	NA	3:05:00
Begin validating docker on each node - Benchmark containers HPL, STREAM. NCCL continued into day 4	NA	3:05:00
Issue mitigation	NA	3:00:00
Documentation issues, benchmark container execution issues	NA	3:00:00
Day 4	(Dell Services work complete)	3:20:00
Validation (continued)	NA	2:20:00
Continuing docker validation - Benchmark containers NCCL	NA	1:00:00
Install Kubernetes	NA	0:50:00
DeepSeek model	NA	0:30:00
Issue mitigation	NA	1:00:00
Rollback and reinstall Kubernetes due to user error	NA	1:00:00

System configuration information

Table 4: Detailed information on the systems we tested.

Server configuration information	Dell™ PowerEdge™ R660 (head node)	Dell PowerEdge R660 (Kubernetes node)	2 x Dell PowerEdge XE9680
BIOS name and version	Dell 2.5.4	Dell 2.5.4	Dell 2.5.4
Non-default BIOS settings	N/A	N/A	Hyper threading disabled, Virtualization disabled
Operating system name and version/build number	Ubuntu 24.03	Ubuntu 24.03	Ubuntu 24.03
Date of last OS updates/patches applied	03/13/2026	03/13/2026	03/13/2026
Power management policy	Performance	Performance	Performance
Processor			
Number of processors	2	2	2
Vendor and model	Intel® Xeon® Gold 5418Y	Intel Xeon Gold 5418Y	Intel Xeon Platinum 8562Y+
Core count (per processor)	24	24	32
Core frequency (GHz)	2.00	2.00	2.80
Memory module(s)			
Total memory in system (GB)	256	256	2,048
Number of memory modules	8	8	32
Vendor and model	Hynix HMCG88AGBRA190N	Hynix HMCG88AGBRA190N	Hynix HMCG94AGBRA177N
Size (GB)	32	32	64
Type	DDR5	DDR5	DDR5
Speed (MHz)	5,600	5,600	5,600
Speed running in the server (MHz)	4,400	4,400	4,400
Storage controller			
Vendor and model	PERC H755 Front	PERC H755 Front	Boss-N1 Monolithic
Cache size	8192 MB	8192 MB	0 MB
Firmware version	52.26.0-5179	52.26.0-5179	2.1.13.2037
Driver version	N/A	N/A	N/A
Local storage - OS			
Number of drives	2	2	2
Drive vendor and model	MICRON MTFDDAK3T8TGA-1B	MICRON MTFDDAK3T8TGA-1B	Dell EC NVMe ISE 7450 RI M.2 80 960GB
Drive size (TB)	3.84	3.84	960
Drive information (speed, interface, type)	6Gbps, SATA, SSD	6Gbps, SATA, SSD	M.2, PCIe Gen4 NVMe, SSD

Server configuration information	Dell™ PowerEdge™ R660 (head node)	Dell PowerEdge R660 (Kubernetes node)	2 x Dell PowerEdge XE9680
Network adapter			
Vendor and model	Broadcom® NetXtreme Gigabit Ethernet (BCM5720)	Broadcom NetXtreme Gigabit Ethernet (BCM5720)	Broadcom NetXtreme Gigabit Ethernet (BCM5720)
Number and type of ports	2 x 1Gb	2 x 1Gb	2 x 1Gb
Firmware version	23.11.4	23.11.4	23.31.1
Network adapter			
Vendor and model	Broadcom Adv. Dual 25Gb Ethernet	Broadcom Adv. Dual 25Gb Ethernet	Mellanox ConnectX-7 MT2910
Number and type of ports	2 x 25Gb	2 x 25Gb	8x IB OSFP 400Gbps
Firmware version	23.11.16.22	23.11.16.22	28.44.2516
GPU			
Vendor and model	NA	NA	NVIDIA® H200
Memory (GB)	NA	NA	141
Firmware version	NA	NA	96.00.DA.00.11
Cooling fans			
Vendor and model	Dell HPR Gold	Dell HPR Gold	Dell HPR Gold
Number of cooling fans	4	4	16
Power Supplies			
Vendor and model	Dell 0FR0KX	Dell 0FR0KX	Dell 01PDR6
Number of power supplies	2	2	6
Wattage of each (W)	1,100	1,100	2,800

Table 5: Detailed configuration information for the network switches we used.

Network switch configuration information	Dell PowerSwitch S4148T-ON	Dell PowerSwitch S5248F-ON	NVIDIA InfiniBand Quantum-2 QM9700
Number and type of ports	48 x 10GbE RJ-45, 4 x 100GbE QSFP28, 2 x 40GbE QSFP+	48 x 25GbE SFP28, 4 x 100GbE QSFP28, 2 x 200GbE QSFP28-DD	64 x NDR, 32 x OSFP
Number and type of ports used in test	8 x 10GbE RJ-45	4 x 25GbE SFP28	16 x NDR

How we tested

Overview

We tested Dell ProDeploy and Rack Integration services against an in-house (DIY) installation and configuration of a Dell AI Factory solution. This methodology shows the steps required by our in-house engineers to reach a functional end-state solution, rather than documenting the full troubleshooting effort we encountered along the way. The procedures are based largely on publicly available NVIDIA documentation, available at https://docs.nvidia.com/dgx-basepod/index.html.

Deploying the head node

Creating the boot loader

1. Open a browser, and go to https://customer.brightcomputing.com/download-iso.
2. Enter the product key (your license key), select Dell EMC from the pulldown menu, and click Download.
3. After downloading the file, create a bootable USB drive from the ISO image you created. If provided the option, select the DD method of writing rather than ISO method.
4. Upon completion of the write operation, eject the USB device.

Installing the head node

1. Insert the USB drive into a USB port on the head node, and boot from the installation media.
2. Choose Start Base Command Manger Graphical Installer, and click Enter.
3. Click Start Installation.
4. Check the I agree box, and click Next.
5. At Kernel Modules, click Next.
6. Verify the hardware, and click Next.
7. At DVD/ISO/USB, click Next.
8. Enter information for time servers, name the servers, and click Next.
9. For HPC Workload Manager, select one, and click Next.
10. Select Type 2, and click Next.
11. Enter the hostname, enter and confirm the administrator password. For the hardware manufacturer, use the pulldown menu, select Dell EMC, and click Next.
12. Set the number of nodes to 2. For the hardware manufacturer, use the pulldown menu, select Dell EMC, and click Next.
13. Select No for the head node, Yes for the compute nodes, and click Next.
14. Enter the iDRAC network information for the compute nodes, and click Next.
15. Enter the production network information, and click Next.
16. Assign an IP address to the target interface, and click Next.
17. Set the IP offset you want to use, and click Next.
18. For Disk layout, select a target device, and click Next.
19. Select One Big Partition, and click Next.
20. At Additional Software, click Next.
21. Review the summary, and click Start.
22. Complete the installation. To Reboot, click Yes..

Configuring the InfiniBand switch

Completing initial configuration

1. Connect the console cable to a laptop and the InfiniBand switch.
2. With your local com port and 115200 baud rate, open a serial port connection.
3. To log in, press Enter.
4. To set up the switch, follow the configuration wizard prompts.

Configuring InfiniBand

1. After setup is complete, at the prompt, type the following:

```
Enable
Config Terminal
ib sm enable
ib sm virt enable
ib smnode 9700-1 create
ib smnode 9700-1 sm-priority 15
Exit
Write mem
```

2. Enter the command:

```
Show ib sm
```

3. Verify the output shows enabled.

Deploying the compute nodes

Setting the MAC address and Boot Loader as TFTP

1. To set the MAC addresses for the nodes, type the following:
   a. `Cmsh > device > use node001 > interfaces > set mac xx:xx:xx:xx:xx:xx`
   b. `Commit`
   c. `Interfaces%: list`
   d. `Interfaces%: use (nic name)`
   e. `(nic name)%: set mac xx:xx:xx:xx:xx:xx`
   f. `commit`
2. Repeat the above for the second compute node (node002).
3. The compute nodes contained Broadcom adapters which requires a shift from HTTPS as boot loader to TFTP. On the head node type:
   a. `Cmsh > device > use node001 > set bootloader TFTP.`
   b. `Commit`
4. Repeat the above for the second compute note (node002).
5. On the front of a compute node, press the power button.
6. Once the system comes up, access the head node and type the following:
   a. `Cmsh > device > use node001 > interfaces.`
7. Copy the names of the IB interfaces. They will be the same names for both nodes.
8. To define the InfiniBand network for the compute nodes, type the following on the head node:
   a. `Cmsh`
   b. `Network`
   c. `Add ibnet`
   d. `Set mtu 9000`
   e. `Set domainname test.lab`
   f. `Set nodebooting no`
   g. `Set lockdownhcpd no`
   h. `Set managementallowed no`
   i. `Set baseaddress x.x.x.x`
   j. `Set dynamicrangestart x.x.x.x`
   k. `Set dynamicrangeend x.x.x.x`
   l. `Set netmaskbits xx`
   m. `Set gateway x.x.x.x`
   n. `Commit`

9. Next type the following:
    a. `Device > use node001 > interfaces`
    b. `add physical {ibnic name} {subnet address} ibnet`
    c. `Commit`
10. Repeat b and c for each InfiniBand interface in your system.
11. Repeat step 9 for node002.
12. Reboot both systems.

Creating and patching the image

Cloning and patching the default image

1. On the head node, type the following:
    a. `Cmsh > Softwareimage > List`
    b. `Clone default-image dell-compute-image`
    c. `Commit`
    d. `Exit`
2. To add missing components to new image, type the following at the bash prompt:
    a. `cm-chroot-sw-image /cm/images/dell-compute-image`
    b. `./{name of dsu binary}`
3. To assign an image to a category, type the following:
    a. `Cmsh > category > add compute`
    b. `Commit`
    c. `Use compute`
    d. `Set softwareimage dell-compute-image`
    e. `Commit`
    f. `exit`
4. To clone the dell-compute-image to a new dell-kubernetes-image, type the following:
    a. `Cmsh > softwareimage > list`
    b. `Clone dell-compute-image dell-kubernetes-image`
    c. `Commit`
5. To assign the new image to the new control-plane category, type the following:
    a. `Category > add control-plane`
    b. `Commit`
    c. `Use control-plane`
    d. `Set softwareimage dell-kubernetes-image`
    e. `Commit`
    f. `Exit`
6. To apply the fresh Dell image, reboot all systems.

Configuring the virtualization environment and validation process

Configuring the Kubernetes control plane

1. To set the MAC addresses for the Kubernetes control plane, type the following:
    a. `Cmsh > device > use control-plane > interfaces > set mac xx:xx:xx:xx:xx:xx`
    b. `Commit`
    c. `Interfaces%: list`
    d. `Interfaces%: use (nic name)`
    e. `(nic name)%: set mac xx:xx:xx:xx:xx:xx`
    f. `Commit`
    g. `Exit`
2. The compute nodes contained Broadcom adapters which requires a shift from HTTPS as boot loader to TFTP. On the head node type:
    a. `Cmsh > device > use control-plane > set bootloader TFTP.`
    b. `Commit`
3. On the front of a compute node, press the power button.

Installing Docker

1. On the head node, type cm-docker-setup.
2. Select deploy, and click OK
3. Accept the default name for the overlay, and click OK.
4. Leave empty, and click OK.
5. Select the compute nodes, and click OK.
6. Accept the default network, and click OK.
7. Leave empty, and click OK.
8. To install the NVIDIA Container Runtime, click Yes, and click OK.
9. Select Save config & deploy, and click OK.
10. To accept the file path for the .conf file, click OK.

Validating Docker

Running HPL Benchmark

1. To log into node00, type `ssh node001`.
2. Type `module load docker`.
3. Type `systemctl start nvidia-fabricmanager.service`.
4. Type `docker info`.
5. To load and run the HPL benchmark container, type the following:
    a. `Docker pull.nvcr.io/nvidia/hpc-benchmarks:25.09`
    b. `MNT="$PWD:/home_pwd"`
    c. `CONT='nvcr.io/nvidia/hpc-benchmarks:25.09'`
    d. `Docker run --gpus all-shm-size=1g -v $MNT $CONT mpirun --bind-to none -np 8 ./hpl.sh --dat/workspace/hpl-linux-x86_64/sample-dat/HPL-8GPUs.dat`
6. Record the results.
7. Repeat the above steps for the second node. Results should not deviate by more than 3%. Our variation was 1.31%

Running the STREAM benchmark

1. To load and run the STREAM benchmark container, type the following:
    a. `Docker run --gpus all --shm-size=1g -v $MNT $CONT ./stream-gpu-test.sh --dt fp32 --t CSAT`
2. Record the results.
3. Repeat the above steps for the second node. Results should not deviate by more than 3%. Our variation was .047%

Running the NCCL benchmark

1. To load and run the NCCL benchmark container, type the following:
   a. `Docker pull ghcr.io/coreweave/nccl-tests:13.1.0-devel-ubuntu24.04-nccl2.29.2-1-2276a5e`
   b. `Docker run --gpus all --ipc=host -it --rm ghcr.io/coreweave/nccl-tests:13.100-devel-ubuntu24.04-nccl2.29.2-1-2276a5e /opt/nccl-tests/build/all_reduce_perf -b 67108864 -e 32G -f 2 -g 8`
2. Record the results.
3. Repeat the above steps for the second node. Results should not deviate by more than 3%. Our variation was 0.0000773%.

Installing Kubernetes

1. On the head node, run cm-kubernetes-setup.
2. Choose deploy, and click OK.
3. Choose the latest version with an asterisk beside it, and click OK.
4. Verify containerd is selected, and click OK.
5. To skip registry mirror server, click OK.
6. Enter the information for the kubernetes cluster, and click OK.
7. To expose to the external network, select Yes, and click OK.
8. Select ibnet, and click OK.
9. Select the control-plane for the kubernetes master node, and click OK.
10. Select the compute nodes for the worker nodes, and click OK.
11. Don't select anything on this screen, and click OK.
12. Select the control-plane node as the etcd node, and click OK.
13. Keep the defaults, and click OK.
14. Select Calico, and click OK.
15. To install the kyverno policy engine, select Yes, and click OK.
16. To decline HA, select No, and click OK.
17. To skip installing policies, select No, and click OK.
18. Select NVIDIA GPU Operator, Prometheus Adapter, Prometheus Adapter Stack, cm-gupyter-kernel-operator, and click OK
19. To be rolled up with defaults, select NVIDIA GPU Operator, Prometheus Adapter, Prometheus Adapter Stack, cm-gupyter-kernel-operator, and click OK.
20. To deploy add-ons, choose Ingress Controller (nginx), Kubernetes Dashboard, Kubernetes Metrics Server, and Kubernetes State Metrics, and click OK.
21. Keep the defaults, and click OK.
22. To skip installing Bright NVIDIA packages, select No, and click OK.
23. Select Yes, and click OK.
24. Accept the empty defaults, and click OK.
25. Select enabled, select default for the local path, and click OK.
26. Accept the defaults, and click OK.
27. Select Save config and Deploy, and click OK.
28. Accept the file path, and click OK. Wait for the installation to complete successfully.

Deploying and validating an AI model

Running DeepSeek

1. To login to node001, type the following:

    a. `ssh node001.`
    b. `module load docker`
    c. `systemctl start nvidia-fabricmanager.service`
    d. `Curl -fsSL https://ollama.com/install.sh | sh`
    e. `ollama pull deepseek-r1:7b`
    f. `ollama run deepseek-r1:7b`

2. Ask the AI a question and verify both the thinking processes and the response.

This project was commissioned by Dell Technologies.

Read the report ▶

Primary contributors

Tech: Craig B., Aaron W.

Writing: Ticia I.

Design: Jared White

PM: Greg Carrero

How we created this report

A PT team, which includes the contributors we've listed and others, created this report and performed the technical work behind it. We used AI to develop the report outline and edit portions of the text.

Principled Technologies®

Facts matter.®